

# 关系嵌入和类协方差度量的小样本图像分类

赵蕊,余玉梅,白梦茹

(云南民族大学 数学与计算机科学学院,云南 昆明 650500)

**摘要:**在小样本图像分类的问题情景下,图像之间具有结构性、辨别性的特征关系还未被重视,图像的特征关系匹配不足,导致特征嵌入不完善,度量效果不明显.为了关注图像特征之间的匹配关系,通过关注单个图像的空间信息及匹配图像间的特征,实现度量嵌入特征的关系,提出基于关系嵌入和类协方差度量的网络结构,通过自相关注意表示和交互相关注意来学习图像内和图像间的特征关系.其中,将单个图像的基本特征表示映射成自相关张量,关注其中的通道空间信息,学习图像张量内结构信息.而对于图像对,计算2个图像表示的互关联信息,并生成共同注意,从而得到图像间具有判别性的关系信息.最后用马氏距离度量特征关系.经过大量实验验证之后,在3个小样本基准数据集中评估,并与大量现有方法进行对比,结果显示网络结构对学习特征间关系具有一定的有效性.

**关键词:**小样本学习;关系嵌入;相关注意;马氏距离

**中图分类号:**TP391.41 **文献标志码:**A **文章编号:**1672-8513(2024)02-0232-10

小样本图像分类目的在于通过少量样本学习新的视觉概念,通过极其少量的支持图像去学习特征表示,并且用适合的方式使得目标类分类准确.然而,模型在学习少量的标识数据时,由于样本的稀疏,以及代表性样本的缺乏,有限的样本之间联系较少,标志特征的匮乏都会成为模型拟合水平表现不佳的原因,这些是小样本图像分类面临的困境.

与此同时,嵌入模型在小样本图像分类的问题上,往往存在过拟合一些不相关的特征,而忽略具有判别性质的特征,有关联的特征之间难以匹配,如何运用特征的相关性便成为嵌入过程中值得关注的问题. Robert 等<sup>[4]</sup>提出在视觉任务中比起纹理特征,应该重视包含视觉底层语义的形状特征, Carl 等<sup>[5]</sup>则认为有监督的学习对特征语义层不太敏感,对同类别的不同图片缺乏区分度.在受一些不匹配特征干扰的情况下,用嵌入模型学习特征关系的全面信息,是有一定难度的.

目前对于小样本图像分类的研究方法有很多分类:基于元学习、基于优化、基于模型等<sup>[6]</sup>. 本文是基于度量的元学习方法,即在元学习结构的形式下,

对支持集样本数据学习嵌入,以此为元知识迁移到查询样本并帮助其分类,其中的两个关键点便是嵌入和度量模块的设计.

因此,本文在训练过程中通过设计嵌入模型<sup>[1-2]</sup>和相似性度量<sup>[3]</sup>减少不相关特征对模型的干扰,提升特征关系匹配的嵌入效果.从考虑相互关系的嵌入模式出发,在相互关联的模式下,学习特征的相关性,既学习图像自身的关注对象是什么,还要学习图像与图像之间在空间位置的匹配特征,在此基础上再形成嵌入,通过关注独立特征关系的马氏距离进一步度量嵌入信息.

本文提出了一个关系嵌入和类协方差度量网络,旨在进一步运用特征表示间的关系.首先运用自相关注意表示,对带有结构信息的自相关张量分别在通道和空间2个维度学习网络应该关注的对象和位置,通过学习强调或抑制信息,有效地帮助网络内的信息流<sup>[7]</sup>得到表示.再通过交互相关注意,将这些表示的信息,通过整合到4维向量中并用4D卷积匹配过滤几何不一致信息,筛选同类别信息,从而形成关系嵌入信息.最终由度量模块对测试的样本

收稿日期:2023-02-01.

基金项目:云南少数民族“百名人才”文化产品O2O电子商务系统建设项目(231600200201048).

作者简介:赵蕊(1999-),女,硕士研究生.主要从事机器学习、小样本学习研究.

通信作者:余玉梅(1965-),女,教授,硕士生导师.主要从事智能计算研究.

与类别之间进行相似度匹配,为了进一步度量特征之间的关系嵌入,用类协方差矩阵估计类别特征的马氏距离<sup>[8]</sup>度量特征关联。

## 1 相关工作

小样本图像分类. 小样本图像分类是提供少量图片样本然后对目标图像识别准确的过程,参与训练的数量具有局限性,一些解决样本数量少的方法被广泛应用到小样本分类中,如数据增强和正则化,对于模型的过拟合有一定的缓解作用,但是却不能够彻底解决数据量少的问题. 目前主流的一些方法主要分为2个分支,一个是元学习分支,基于度量的方法,通过学习一个嵌入模型将图像信息映射到度量空间,再由距离函数测量图像对之间的距离,根据距离远近判断类别<sup>[9-10]</sup>. 基于优化的方法,在面对数据量少而导致模型欠拟合严重的问题下,通过优化模型当中的参数,使其具有良好的泛化能力<sup>[11]</sup>. 基于模型的方法则是通过一些外部存储器组件等减轻低数据状态下的训练负担,构建快速泛化的模型<sup>[6]</sup>. 另一个分支便是迁移学习,通过转移已经学习过的相关任务中的知识,进而来改进新任务中的表现. 本文属于元学习分支下基于度量的方法,基本思路是:给定训练的数据,从训练的数据中学习得到嵌入空间,通过插值或者外推特征,将训练数据上训练得的嵌入模型应用到不可见目标类上<sup>[12-13]</sup>,再由固定的距离度量或者是可学习距离的网络度量出样本间相似度,同类的距离近,相似度就高,反之相似度低。

自相关,互相关. 由Lowe等<sup>[14]</sup>提出局部自相似描述子,通过描述尺度空间中的关键位置、关键点并生成特定的特征向量. 自相关通过测量局部的相似程度来比对图像结构布局<sup>[15]</sup>,通常被用作描述符,被深度神经网络用来学习有效表示,本文受Robert等<sup>[4]</sup>和ZHENG等<sup>[16]</sup>对于空间结构特征看法的启发,引入了一个自相关注意模块,与之前的自相关处理模式不同的是:使用了通道相关,保留了特征向量的丰富语义,并使用通道和空间注意,为图像识别提供目标对象空间尺度更聚集的特征信息,利用这些信息作为自相关张量,进而学习训练样本对象. 互相关的概念广泛用于通信信号领域,用于2个有关联关系的变量或者信号上,通过相关函数测量相似度程度. 在图像视觉领域,支持或查询集中的测试图像由于未经训练,因此特征很难关注到目标对象. 也可以理解为,对于包含多个对象的测试图像,特征会更

关注已经在训练集中出现过的参与训练的对象,而忽略未经训练的对象,因此CAN<sup>[10]</sup>提出1种通过考虑可见类和不可见类的样本中最相关的区域去提高特征分辨力的方法. DN4<sup>[2]</sup>中,图像的局部特征在数量较少的时候使用图像对图像的度量不如图像到类的度量,所以提出基于局部描述符的图像到类的度量去代替基于特征的度量. 而本文的互相关是关注特征之间的几何关系,利用4D卷积提炼互相关张量,过滤掉几何不一致的相关信息,由此获得可靠的注意信息,减轻训练过程中由于图像外观变化大导致相关性不强的影响。

度量. 通常在基于元学习设定的小样本分类的任务下,训练类内的数量较少,会出现分类器优化困难等问题. 度量的目的是学习同类样本距离近且不同类样本距离远的度量空间<sup>[18]</sup>. 度量方式的选择可以根据样本特征的关注点去选择,对于独立特征的度量方法可以选择欧氏距离、余弦距离等. 但是这些方法会忽略特征之间的相关性,而马氏距离计算向量间的距离时,考虑到特征间的相关性,因此本文考虑学习特征之间相关性的马氏度量,进一步从度量函数关注特征的相关性。

## 2 小样本图像分类任务描述

通常情况下,小样本图像分类,使用的是任务训练机制,一个任务 $T$ 中包含支持集(support set)和查询集(query set),每个支持集和查询集以 $N$ -way  $K$ -shot的形式分配,即每个集有 $N$ 个类别,每个类别有 $K$ 个样本. 在小样本分类时,从训练类 $C_{train}$ 中选取训练数据 $D_{train}$ 训练优化模型,在测试类 $C_{test}$ 中,选取测试数据 $D_{test}$ 评估,其中 $C_{train}$ 和 $C_{test}$ 不相交,训练数据的标签空间与测试数据的标签空间也不相交,且均按照任务机制分组. 训练的时候,训练数据 $D_{train}$ 迭代采样每个任务分组,让模型从支持集和查询集中学习映射,再用映射去测试 $D_{test}$ 中的查询集属于支持集的哪个类别。

## 3 模型介绍

关系嵌入和类协方差度量网络(RECMN: relation embeddings and covariance measures network),可以分为2个部分,第1部分就是关于嵌入模块的构建,其中包含关注图像自身的自相关注意表示模块和关注图像对之间关系的交互相关注意模块. 第2部分是度量嵌入向量的度量模块,根据不同类别的协方差是不一致的,考虑使用类别相关的马氏距离度量. 整个网络架构图如图1所示。

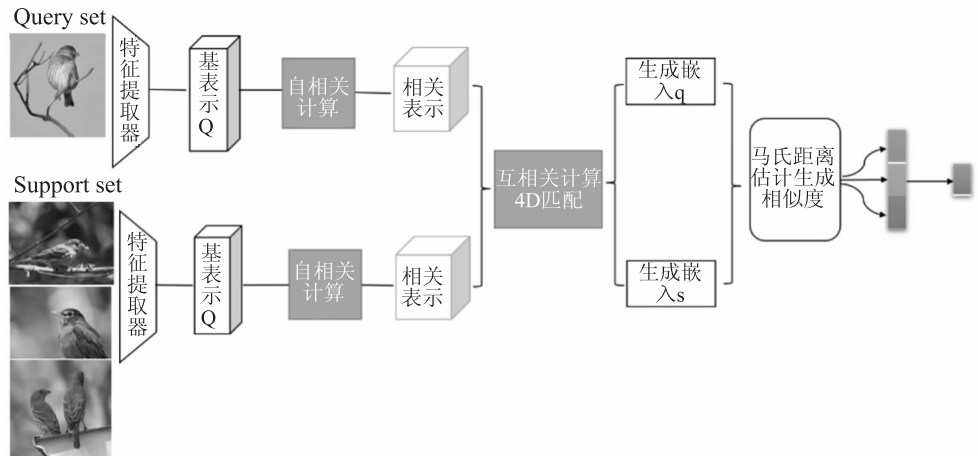


图 1 RECMN 网络架构图

### 3.1 嵌入模块概述

嵌入模块由 2 个模块构成,自相关注意表示 (SCAR)和交互相关注意(CCA)模块.首先,用特征提取器对查询和支持图像提取基础表示,然后经过 SCAR 模块,即基础表示先通过自相关计算形成张量,再由带有通道空间注意机制的卷积块,将基础表示转变为自相关表示 F. 其次 CCA 模块对自相关表示学习,分别生成支持和查询的共同注意映射  $A_q, A_s$ , 对自相关表示 F 加权映射,最终得到图像嵌入 q 和 s.

### 3.2 自相关注意表示(SCAR)

自相关注意表示 (SCAR; self - correlational at-

tentional representation) 模块:转换自相关张量为带有通道空间注意映射的结构相关的信息,进而给后续的特征间关系匹配提供输入,由自相关计算、自相关表示构成.整个过程为:从特征提取器得到查询图片和支持图片的基表示,通过自相关计算得到张量,后使用通道空间注意机制关注张量中关键区域信息,并用卷积块降维以便被添加到基表示上,最终形成关注自身信息的自相关表示对,为后续匹配特征关系的 CCA 模块提供输入. SCAR 模块的实现如图 2 所示.

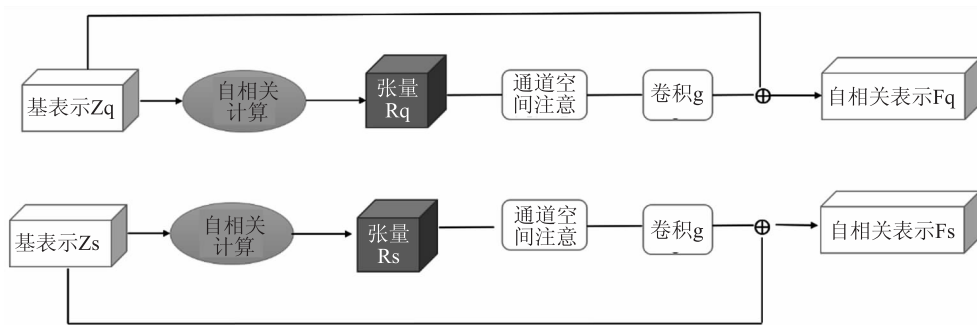


图 2 SCAR 模块构图

#### 3.2.1 自相关计算

对特征提取器得到的基表示  $z \in R^{H \times W \times C}$  ( $H \times W \times C$  代表基表示的维度),计算每个像素点的  $C$  维向量及其  $x$  轴,  $y$  轴邻域向量的 Hadamard 积,将其转为自相关张量  $R \in R^{H \times W \times U \times V \times C}$  ( $H \times W \times U \times V \times C$  代表张量的维度),由式(1)可以得出:

$$R(x, p) = \frac{z(x)}{\|z(x)\|} \odot \frac{z(x+p)}{\|z(x+p)\|}. \quad (1)$$

其中的  $p$  表示像素点  $x \in [1, H] \times [1, W]$  的空间位置,即  $p \in [-du, du] \times [-dv, dv]$ ,沿着  $x$  轴和  $y$  轴

方向,像素点的空间范围为  $2du + 1 = U$  和  $2dv + 1 = V, U, V$  均为 5. 这样保留空间位置的形式,能够揭示图像的结构模式,进而保留特征向量的丰富语义便于学习关系.

#### 3.2.2 自相关表示

为了进一步分析自相关张量  $R$  中的空间结构关键信息,分别对张量  $R$  在空间和通道两个维度提取特征,通过卷积运算混合跨通道和空间得到张量  $R'$ 如图 3,然后再沿  $U \times V$  维应用一系列 2 维卷积.

在对通道和空间维度用卷积提取特征时, 对于空间信息的聚合, 普遍采用的是平均池化. Zhou 等<sup>[19]</sup>建议使用平均池化来有效地学习目标对象的范围, WOO 等<sup>[7]</sup>则认为最大池化收集了关于独特对象特征的重要线索, 以推断更精细的通道式注意力. 而本文中 SCAR 模块结合 2 种池化, 对自相关张量  $R$  即使用了平均池化又使用最大池化最后输出张量  $R'$ .

分别在通道和空间维度处理完特征后, 沿  $U \times V$

维再进行卷积块  $g(\cdot)$  学习自相关模式, 即使用通道尺寸缩小的点向卷积层、变换的  $3 \times 3$  卷积层和恢复通道的点向卷积层组成卷积块逐步聚合局部相关模式. 最终将空间维数  $U \times V$  降低到  $1 \times 1$ , 输出卷积后信息与基表示就能保持一个维度结合, 为了将学习得到的注意映射信息聚合到基表示上, 需要式 (2) 计算得自相关表示  $F \in R^{H \times W \times C}$ ,  $Z$  为基表示:

$$F = g(R') + Z. \quad (2)$$

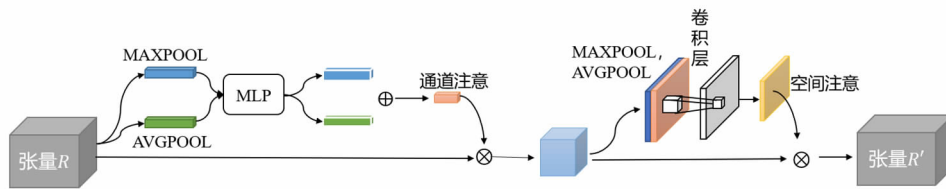


图 3 对  $R$  跨通道和空间提取信息

### 3.3 交互相关注意 (CCA)

交互相关注意模块 (CCA: cross-correlational attention): 对自相关表示信息对进行关系匹配, 并将空间结构不一致信息剔除. 模块由互相关计算, 卷积匹配, 互注意加权组成. 整个过程为: 将 SCAR 模块

学习完的自相关表示对作为 CCA 模块的输入, 进行互相关计算, 得到相关张量, 并用 4D 卷积匹配筛选相关张量中几何一致的特征, 得到筛选后的匹配张量用来计算注意映射, 最终完成对自相关表示对的关系嵌入. 具体构架如图 4 所示.

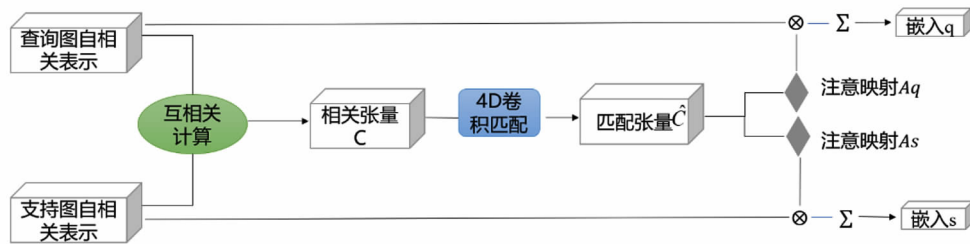


图 4 CCA 模块构图

#### 3.3.1 互相关计算

首先对查询和支持自相关表示 ( $F_q$  和  $F_s \in R^{H \times W \times C}$ ) 用点向卷积层, 将高维特征先降维得到表示  $\hat{F}_q$  和  $\hat{F}_s$ , 便于构建 4 维张量, 利用特征间的余弦相似度对点向卷积后的表示对构造四维张量  $C \in R^{H \times W \times H \times W}$  如式 (3), 便于后续 2 种表示之间进行特征筛选,  $x$  表示特征图上的空间位置,  $sim(\cdot)$  表示两个特征之间的余弦相似度:

$$C(x_q, x_s) = sim(\hat{F}_q(x_q), \hat{F}_s(x_s)). \quad (3)$$

#### 3.3.2 卷积匹配

由于张量  $C$  包含了所有成对匹配的分对, 为了消除不匹配信息进行特征筛选, 采用卷积匹配过程. 在 4D 空间中, 几何模式与输入图像中的平移是

等变的, 如果图像被平移, 匹配模式在 4D 空间也会发生等量的平移<sup>[17]</sup>. 4D 空间位置不影响匹配信息. 因此使用 4D 卷积处理相关图, 过滤几何不一致的相关, 实现特征空间匹配的作用.

匹配过程通过带匹配核的 4D 卷积处理张量  $C$ . 卷积匹配  $h(\cdot)$  由 2 个 4D ( $3 \times 3 \times 3 \times 3$ ) 卷积层组成, 卷积层之间插入批处理归一化和 ReLu, 第一个卷积层产生具有匹配核的相关张量, 由输入通道 1 增加到  $C_1$ , 通过消融实验最终设置为 16; 第二个卷积层将第一个卷积层得到的信息聚合为通道数为 1 的单个 4D 相关张量  $\hat{C} = h(C) \in R^{H \times W \times H \times W}$ . 4D 卷积的维度变化示意图如图 5 所示.

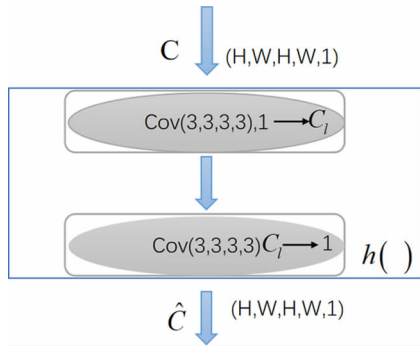


图5 卷积匹配维度示意图

### 3.3.3 互注意加权

互注意加权从卷积匹配的张量 $\hat{C}$ 中,生成查询对支持的注意映射 $A_q$ 和支持对查询的注意映射 $A_s$ ,注意映射表示查询与支持之间的相关权值,用于进行关系嵌入。

$$A_q(x_q) = \frac{1}{HW} \sum_{x_s} \frac{\frac{\exp(\hat{C}(x_q, x_s))}{\tau}}{\sum_{x'_q} \frac{\exp(\hat{C}(x'_q, x_s))}{\tau}}. \quad (4)$$

式(4)中 $\tau$ 是个参数、 $HW$ 表示张量 $\hat{C}$ 的维度,可以将其理解为将匹配分数变换为注意映射,即计算查询图像中某个位置 $x_q$ 与支持图像中某个位置 $x_s$ 匹配的平均概率。映射 $A_s$ 计算方法也是相同的,仅对式(4)中的查询图和支持图交换位置后计算出注意映射。

为了获得最终的嵌入信息,要将注意映射分别添加到自相关表示对。查询自相关表示 $F_q \in R^{H \times W \times C}$ 乘以查询映射 $A_q$ 得到查询图像的关系嵌入 $q$ :

$$q = \sum_{x_q} A_q(x_q) F_q(x_q). \quad (5)$$

类似对支持自相关表示进行映射得到嵌入 $s$ ,最终实现支持图像的关系嵌入:

$$s = \sum_{x_s} A_s(x_s) F_s(x_s). \quad (6)$$

### 3.4 度量

由于小样本分类任务特点是训练数据较少,会加大模型过拟合的风险,降低模型的泛化性能。因此在特征嵌入后,非参数化的分类模型相对降低了特征提取器的训练难度,同时,也使得模型的结构更加灵活,可以快速地识别新的类别<sup>[20]</sup>。参考BATENI等<sup>[8]</sup>的方法,关注特定任务的独立特征间关系,用特定于任务和类别的协方差矩阵,度量特征嵌入的马氏距离。不同之处在于度量的向量是带有关系匹配信息的,可以进一步度量特征关系。

通过使用固定的马氏距离 $d_k$ 如式(7)进行度量,其中 $Q_k^T$ 是任务 $T$ 的类协方差矩阵,其中 $k$ 表示类别数。

$$d_k(x, y) = \frac{1}{2}(x - y)(Q_k^T)^{-1}(x - y). \quad (7)$$

要求类协方差矩阵,需要对类内协方差矩阵和类间协方差矩阵计算,用式(8)对任务内的所有类计算类内协方差矩阵 $M_k^T$ ,而类间协方差矩阵 $M^T$ 对支持集中所有样本进行计算,其计算方法和类内协方差矩阵类似。其中 $\mu_k$ 表示类均值, $S_k^T$ 表示支持集的任务 $T$ 中有 $k$ 个标签的类别数。 $f_\theta^T(x_i)$ 表示 $x_i$ 查询集样本的特征嵌入, $f$ 表示嵌入模型。最后再用softmax函数评估样本相似程度。

$$M_k^T = \frac{1}{|S_k^T| - 1} \sum_{(x_i, y_i) \in S_k^T} (f_\theta^T(x_i) - \mu_k)(f_\theta^T(x_i) - \mu_k)^T. \quad (8)$$

要求类协方差矩阵还需要正则化类内和类间协方差矩阵。由于支持集的样本数量太少,比特征空间维度小的多,为了保证可逆,使用式(9)正则化估计器对每个任务的类内协方差矩阵 $M_k^T$ 和任务间的类间协方差矩阵 $M^T$ 进行正则化,并通过矩阵缩放系数 $\beta$ 和比例系数 $\lambda_k^T$ 加权组合形成类协方差矩阵 $Q_k^T$ ,其中 $I$ 为单位矩阵。而式(9)中 $\lambda_k^T$ 是比例系数,可以由式子(10)计算得到。

$$Q_k^T = \lambda_k^T M_k^T + (1 - \lambda_k^T) M^T + \beta I. \quad (9)$$

$$\lambda_k^T = \frac{|S_k^T|}{(|S_k^T| + 1)}. \quad (10)$$

### 3.5 训练方式

大多方法是按照2阶段,先预训练,后情景训练的训练方案<sup>[21-22]</sup>,然而这种方式比较耗时,本文则选择比较省时的单阶段训练的方式<sup>[23]</sup>。训练阶段由两种损失来联合训练所提出的模型,基于锚点的分类损失 $L_{\text{anchor}}$ 以及基于度量的分类损失 $L_{\text{metric}}$ 。

基于锚点的分类损失通过对查询集的基表示 $Z_q$ 计算,指导损失模型对训练集 $C_{\text{train}}$ 中的查询图像类别进行正确分类,由式(11)计算得到, $w_c^T$ 和 $b_c$ 是全连接层的参数:

$$L_{\text{anchor}} = -\log \frac{\exp(w_c^T z_q + b_c)}{\sum_{c'=1}^{|C_{\text{train}}|} \exp(w_{c'}^T z_q + b_{c'})}. \quad (11)$$

基于度量的损失是取训练集 $D_{\text{train}}$ 并将其分为 $T$ 个任务,每个任务由支持集和查询集组成。接着将划分好的任务输入到嵌入模块中,计算支持集的原型 $s$ 。计算查询集中各个样本 $q_k$ 与支持集类原型间的

马氏距离  $d_k$ , 利用式(12) 计算  $L_{\text{metric}}$  损失, 其中  $N$  表示类别数,  $k$  表示样本数.

$$L_{\text{metric}} = -\log \frac{\exp(d_k(\bar{s}, q_k))}{\sum_{n=1}^N \exp(d_k(\bar{s}, q_k))}. \quad (12)$$

最终损失由两部分组成, 如式(13), 其中  $\lambda$  为平衡损失项的超参数:

$$L = L_{\text{anchor}} + \lambda L_{\text{metric}}. \quad (13)$$

## 4 实验

### 4.1 数据集

为了验证网络有效性, 分别在3个小样本分类基准上进行了实验: miniImagenet、CUB-200-2011<sup>[24]</sup> 和 CIFAR-FS. 并对3个基准数据集都进行了训练、验证、测试分组. miniImageNet<sup>[1]</sup> 是 ImageNet 的一个子集, 共有 60 000 张图片, 由 100 个类组成, 每个类有 600 个样本, 将其中 64 个类作为训练集、16 个类作为验证集、20 个类作为测试集. CUB-200-2011 (CUB) 是一个包含 200 个种类, 11 788 张图像的鸟类细粒度分类数据集, 分别将 100、50、50 个对象类别用于训练、验证、测试. CIFAR-FS<sup>[25]</sup> 全称为 CIFAR-100 小样本数据集, 是来自于 CIFAR-100 数据集中的一个子集, 共包含 100 个类别, 将 64、16、20 个类别分别组成训练、验证、测试集, 每个类别 600 张图像, 合计 60 000 张  $32 \times 32$  像素的彩色图像.

### 4.2 实验细节

网络采用 ResNet12<sup>[26]</sup> 作为特征提取网络. 对特征提取网络输入空间大小为  $84 \times 84$  的图像, 特征提取后的基表示  $Z \in R^{5 \times 5 \times 640}$  为 SCAR 模块的输入. SCAR 模块中自相关计算中  $U$ 、 $V$  维度参数均为 5. 在 CCA 模块里, 采用 4D 卷积<sup>[27]</sup>, 为了减少内存量和提高效率<sup>[17]</sup> 其核大小设置为  $3 \times 3 \times 3 \times 3$ . 互注意加权模块计算注意映射时参数  $\tau$  的设置, 在 miniImagenet, CIFAR-FS, CUB 中分别设置为 5, 5, 2. 对于

$N$ -way  $k$ -shot, 分别在类别数  $N$  为 5,  $K$  为 1 和 5 个样本数量上进行训练. 计算损失时用到的超参数  $\lambda$ , 在 miniImagenet, CIFAR-FS, CUB 中分别设置为 0.25, 0.5, 1.5.

### 4.3 实验结果

为体现本文所提出网络结构的有效性, 分别在 2 个基准数据集和一个细粒度数据集上进行实验, 结果如表 1、2 所示, 与往年方法的准确性进行比较, 所有方法的特征提取网络均使用了 Resnet12.

表 1 显示的是在 miniImagenet 和 CIFAR-FS 数据集上对 5-way 1-shot 和 5-shot 任务进行实验, 可以看出本文的网络结构 (RECMN), 提升了小样本图像分类的性能. 与所列方法的最佳结果对比, 在 miniImagenet 数据集, 1-shot 任务中, 比 Cosine classifier、ProtoNet、MetaOptNet、MatchNet、CAN、DeepEMD、FEAT、Tian 等、DSN、Kim 等分别提高了 19.24%、12.28%、12.03%、11.59%、10.82%、8.76%、7.89%、9.85%、12.03%、9.59%. 5-shot 任务中, 分别提高 12.08%、8.73%、10.63%、13.27%、9.82%、6.85%、7.21%、7.12%、10.46%、6.56%. 在 CIFAR-FS 数据集, 1-shot 的任务中比 ProtoNet、MetaOptNet、DeepEMD、Tian et al.、DSN、Kim et al. 分别提高了 14.99%、14.59%、12.61%、13.29%、14.89%、13.68%. 5-shot 任务中分别提高 10.35%、9.55%、6.93%、6.95%、8.75%、8.36%.

表 2 显示的是在细粒度数据集 CUB 上实验后与一些先进方法的对比结果, 可以发现在 5-way 1-shot 任务中分别比 ProtoNet、Cosine classifier、MatchNet、FEAT、DeepEMD 提高了 17.13%、15.92%、11.35%、9.95%、6.08%. 在 5-shot 任务中分别提高了 15.46%、13.21%、12.88%、12.19%、8.98%. 可见无论在细粒度还是标准数据集上, 本文所提出的网络都有改进.

表 1 miniImagenet、CIFAR-FS 数据集 5-way 的分类结果

%

网络	miniImagenet		CIFAR-FS	
	1-shot	5-shot	1-shot	5-shot
Cosine classifier <sup>[28]</sup>	55.43 ± 0.81	77.18 ± 0.61	—	—
ProtoNet <sup>[31]</sup>	62.39 ± 0.21	80.53 ± 0.14	72.2 ± 0.7	83.5 ± 0.5
MetaOptNet <sup>[32]</sup>	62.64 ± 0.82	78.63 ± 0.46	72.6 ± 0.7	84.3 ± 0.5
MatchNet <sup>[1]</sup>	63.08 ± 0.80	75.99 ± 0.60	—	—
CAN <sup>[10]</sup>	63.85 ± 0.48	79.44 ± 0.34	—	—

续表 1

网络	miniImagenet		CIFAR - FS	
	1 - shot	5 - shot	1 - shot	5 - shot
DeepEMD <sup>[35]</sup>	65.91 ± 0.82	82.41 ± 0.56	74.58 ± 0.29	86.92 ± 0.41
FEAT <sup>[36]</sup>	66.78 ± 0.20	82.05 ± 0.14	-	-
Tian et al. <sup>[38]</sup>	64.82 ± 0.60	82.14 ± 0.43	73.90 ± 0.80	86.90 ± 0.50
DSN <sup>[37]</sup>	62.64 ± 0.66	78.8 ± 0.45	72.30 ± 0.80	85.10 ± 0.60
Kim et al. <sup>[33]</sup>	65.08 ± 0.86	82.70 ± 0.54	73.51 ± 0.92	85.49 ± 0.68
(Our) RECMN	74.67 ± 0.44	89.26 ± 0.22	87.19 ± 0.27	93.85 ± 0.17

表 2 CUB 数据集 5 - way 的分类结果

网络	特征提取器	1 - shot/%	5 - shot/%
ProtoNet[31]	Resnet12	66.09 ± 0.92	82.50 ± 0.58
Cosine classifier <sup>[28]</sup>	Resnet12	67.30 ± 0.86	84.75 ± 0.60
MatchNet <sup>[1]</sup>	Resnet12	71.87 ± 0.85	85.08 ± 0.57
FEAT[36]	Resnet12	73.27 ± 0.22	85.77 ± 0.14
DeepEMD <sup>[35]</sup>	Resnet12	77.14 ± 0.29	88.98 ± 0.49
(Our) RECMN	Resnet12	83.22 ± 0.37	97.96 ± 0.09

由表 1 和表 2 的实验结果可见,无论在细粒度还是标准数据集上,RECMN 方法均具有一定有效性,即在小样本图像分类中,通过交叉式学习图像关系的特征嵌入能够有效地促进学习特征间的关系,提升特征匹配的性能,同时,通过引入学习独立特征间关系的马氏距离,进一步增强了对交叉式特征嵌入关系的度量作用.由此可知,学习图像特征间关系对于小样本图像分类的嵌入模块和度量模块均很重要.

#### 4.4 消融实验

SCAR 模块的注意力机制消融实验.为了验证通道和空间注意力对 SCAR 模块提取特征空间信息的作用,在添加位置进行了消融实验.分别在 SCAR

模块的 3 个位置做了消融实验.

以在 CUB 数据集上实验,5 - way 1 - shot 任务为例,由表 3 可看出,注意力机制在自相关计算的张量 R 之后添加比在自相关表示的卷积块  $g(\cdot)$  后添加和在 2 个位置均添加的效果分别高出 0.42%、0.67%,在 2 个位置处均添加的效果是比较差的,仅 79.05%.因此,最终选择在自相关计算得到的张量 R 之后添加.可以推断自相关计算之后的张量包含的原始图像信息是最丰富的,张量中的结构信息对于关注通道空间信息的注意力设计具有帮助效果,能注意到更多初始的关系信息,有利于网络进一步的学习特征.

表 3 通道空间注意的位置影响

添加位置			数据集
自相关计算的张量 R 后	自相关表示的卷积层 $g(\cdot)$ 后	R 与 $g(\cdot)$ 后	CUB(5 - way 1 - shot)
√			79.72 ± 0.43
	√		79.30
		√	79.05 ± 0.44

CCA 模块的通道变化设计消融实验.为了验证卷积匹配中通道数对模块的影响,在排除 SCAR 模块的影响下,对 CCA 模块中卷积匹配的 2 个卷积层的输入、中间、输出通道数进行消融实验,可以从图 5 中看到卷积匹配的通道变化图.

设置方案(a)通过消融 CCA 模块中所有可学习的参数来检测非参数基线,用基表示  $Z_q$  和  $Z_s$  作为嵌入信息直接参与关系度量.方案(b)设置初始的 3 层通道数量,验证了通道维度的初始设置有助于捕捉到互相关信息.方案(c)为 CCA 模块互相关计算中,构建 4 维

向量  $C \in R^{H \times W \times H \times W}$  时使用哈达玛积而不是使用余弦相似度,保持维度的不变性,但耗费的时间更多,需要学习的参数过多,效果不显著. 方案(d)则使用余弦相似度构建四维向量,将卷积匹配输入、中间、输出通道设置为 1、16、1,卷积匹配通道变化实验结果由表 4 所示,

表 4 CCA 模块卷积匹配通道数设计的影响

编号	卷积匹配中通道变化(输入层 - 中间层 - 输出层)	miniImagenet(5 - way 1 - shot)/%	CUB(5 - way 1 - shot)/%
(a)	不设置	65.73 ± 0.14	77.75 ± 0.54
(b)	1 - > 1 - > 1	65.75 ± 0.56	78.05 ± 0.45
(c)	64 - > 16 - > 1	66.18 ± 0.12	78.10 ± 0.87
(d)	1 - > 16 - > 1	65.90 ± 0.24	78.49 ± 0.76

度量方式的对比实验. 能够关注独立特征的度量方式有许多种,余弦距离度量和马氏距离度量都可以做到. 因此,为了验证学习独立特征间关系的马氏距离能够增强所提出的特征嵌入模式的效果,通过比较关注独立特征不同度量方式的影响,来探讨这些方法对交叉关系嵌入模型的差异性. 这样做进一步证实了最终选择的度量方法对于增强交叉关系嵌入的有效性. 因此在度量方式的选择上,对余弦距离度量方式和关注特征之间关系的马氏距离度量方式进行了对比实验.

对比的实验结果可从表 5 和表 6 中看出:在 5 - way 1 - shot 任务中,CUB 数据集上由余弦距离作为度量方式的 79.49% 提升到马氏距离作为度量方式的 83.22%,miniImagenet 数据集上由 67.6% 提高到 74.67%,CIFAR - FS 数据集上由 74.51% 提升到 87.19%. 同样的,在 5 - way 5 - shot 的任务中,CUB 数据集上由余弦距离作为度量方式的 91.11% 提升到马氏距离作为度量方式的 97.96%,miniImagenet 数据集上由 82.58% 提升到 89.26%,CIFAR - FS 数据集上由 86.6% 提升到 93.85%.

表 5 余弦距离度量在 3 个基准数据集上的表现 %

数据集	5 - way 1 - shot	5 - way 5 - shot
CUB	79.49 ± 0.21	91.11 ± 0.71
miniImagenet	67.60 ± 0.56	82.58 ± 0.27
CIFAR - FS	74.51 ± 0.8	86.60 ± 0.54

表 6 马氏距离度量在 3 个基准数据集上的表现 %

数据集	5 - way 1 - shot	5 - way 5 - shot
CUB	83.22 ± 0.37	97.96 ± 0.09
miniImagenet	74.67 ± 0.44	89.26 ± 0.22
CIFAR - FS	87.19 ± 0.27	93.85 ± 0.17

在 miniImagenet、CUB 两个数据集上进行 5 - way 1 - shot 实验,miniImagenet 实验得到的数据并不是最佳的. 但是与最佳结果差异不大,同样设置在 CUB 上是最佳效果,因此考虑到计算时长和效果差异不大的因素,选用方案(d)作为卷积匹配的通道变化方案.

由此可见,能够学习独立特征之间关系的马氏距离度量效果对带有关系嵌入的信息度量效果表现更佳. 由于每个数据集之间的样本分布是不一致的,而数据集中的类内和类间也存在差异,因此能够学习样本协方差矩阵差异的马氏距离对关系嵌入的信息的度量效果带来很大的帮助.

### 5 结语

提出 1 种新的基于度量的小样本图像分类的网络结构(RECMN),利用自相关注意表示筛选关注图像自身的空间信息,结合交互相关注意匹配图像对之间的几何一致信息,将其转换为注意映射并完成关系嵌入,再由学习样本协方差矩阵差异的马氏距离度量关系嵌入信息,最终在两个标准数据集和一个细粒度数据集上均达到比较好的提升效果,对学习特征间关系的研究在图片分类的应用上带来了非常大的性能提升. 但仍然还存在改进空间,在未来工作中,还可以从样本分布对特征关系做进一步探究.

### 参考文献:

[1] VINYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning[J]. Advances in neural information processing systems, 2016, 29.

[2] LI W, WANG L, XU J, et al. Revisiting local descriptor based image - to - class measure for few - shot learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7260 - 7268.

[3] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: Relation network for few - shot learning [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1199 - 1208.

[4] GEIRHOS R, RUBISCH P, MICHAELIS C, et al. ImageNet - trained CNNs are biased towards texture; increasing

- shape bias improves accuracy and robustness [J]. arXiv preprint arXiv:1811.12231, 2018.
- [5] DOERSCH C, GUPTA A, ZISSERMAN A. Crosstransformers: spatially-aware few-shot transfer [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 21981–21993.
- [6] PARNAMI A, LEE M. Learning from few examples: A summary of approaches to few-shot learning [J]. arXiv preprint arXiv:2203.04291, 2022.
- [7] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]//*Proceedings of the European conference on computer vision (ECCV)*. 2018:3–19.
- [8] BATENI P, GOYAL R, MASRANI V, et al. Improved few-shot visual classification [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:14493–14502.
- [9] ALLEN K, SHELHAMER E, SHIN H, et al. Infinite mixture prototypes for few-shot learning [C]//*International Conference on Machine Learning*. PMLR, 2019:232–241.
- [10] HOU R, CHANG H, MA B, et al. Cross attention network for few-shot classification [J]. *Advances in Neural Information Processing Systems*, 2019, 32.
- [11] WANG Y, YAO Q, KWOK J T, et al. Generalizing from a few examples: A survey on few-shot learning [J]. *ACM computing surveys (csur)*, 2020, 53(3):1–34.
- [12] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015:815–823.
- [13] TENENBAUM J. Mapping a manifold of perceptual observations [J]. *Advances in neural information processing systems*, 1997, 10.
- [14] LOWE D G. Object recognition from local scale-invariant features [C]//*Proceedings of the seventh IEEE international conference on computer vision*. Ieee, 1999, 2:1150–1157.
- [15] SHECHTMAN E, IRANI M. Matching local self-similarities across images and videos [C]//*2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007:1–8.
- [16] ZHENG C, CHAM T J, CAI J. The spatially-correlative loss for various image translation tasks [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021:16407–16417.
- [17] ROCCO I, CIMPOI M, ARANDJELOVIĆ R, et al. Neighbourhood consensus networks [J]. *Advances in neural information processing systems*, 2018, 31.
- [18] ZHANG X, YU F X, KARAMAN S, et al. Heated-up softmax embedding [J]. arXiv preprint arXiv:1809.04157, 2018.
- [19] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:2921–2929.
- [20] 葛轶洲, 刘恒, 王言, 等. 小样本困境下的深度学习图像识别综述 [J]. *软件学报*, 2021, 33(1):193–210.
- [21] WERTHEIMER D, TANG L, HARIHARAN B. Few-shot classification with feature map reconstruction networks [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021:8012–8021.
- [22] YE H J, HU H, ZHAN D C, et al. Few-shot learning via embedding adaptation with set-to-set functions [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020:8808–8817.
- [23] ORESHKIN B, RODRÍGUEZ LÓPEZ P, LACOSTE A. Tadam: Task dependent adaptive metric for improved few-shot learning [J]. *Advances in neural information processing systems*, 2018, 31.
- [24] WAH C, BRANSON S, WELINDER P, et al. The caltech-ucsd birds-200-2011 dataset [J]. 2011.
- [25] BERTINETTO L, HENRIQUES J F, TORR P H S, et al. Meta-learning with differentiable closed-form solvers [J]. arXiv preprint arXiv:1805.08136, 2018.
- [26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770–778.
- [27] YANG G, RAMANAN D. Volumetric correspondence networks for optical flow [J]. *Advances in neural information processing systems*, 2019, 32.
- [28] CHEN W Y, LIU Y C, KIRA Z, et al. A closer look at few-shot classification [J]. arXiv preprint arXiv:1904.04232, 2019.
- [29] LIU Y, LEE J, PARK M, et al. Learning to propagate labels: Transductive propagation network for few-shot learning [J]. arXiv preprint arXiv:1805.10002, 2018.
- [30] SUN Q, LIU Y, CHUA T S, et al. Meta-transfer learning for few-shot learning [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019:403–412.
- [31] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [J]. *Advances in neural information processing systems*, 2017, 30.
- [32] LEE K, MAJI S, RAVICHANDRAN A, et al. Meta-learning with differentiable convex optimization [C]//*Proceedings of the IEEE/CVF conference on computer vision and*

- pattern recognition. 2019;10657 – 10665.
- [33] KIM J, KIM H, KIM G. Model – agnostic boundary – adversarial sampling for test – time generalization in few – shot learning [ C ] // Computer Vision – ECCV 2020; 16th European Conference, Glasgow, UK, August 23 – 28, 2020, Proceedings, Part I 16. Springer International Publishing, 2020; 599 – 617.
- [34] LI H, EIGEN D, DODGE S, et al. Finding task – relevant features for few – shot learning by category traversal [ C ] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019; 1 – 10.
- [35] ZHANG C, CAI Y, LIN G, et al. Deepemd: Few – shot image classification with differentiable earth mover’s distance and structured classifiers [ C ] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020; 12203 – 12213.
- [36] YE H J, HU H, ZHAN D C, et al. Few – shot learning via embedding adaptation with set – to – set functions [ C ] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020; 8808 – 8817.
- [37] SIMON C, KONIUSZ P, NOCK R, et al. Adaptive subspaces for few – shot learning [ C ] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020; 4136 – 4145.
- [38] TIAN Y, WANG Y, KRISHNAN D, et al. Rethinking few – shot image classification; a good embedding is all you need? [ C ] // Computer Vision – ECCV 2020; 16th European Conference, Glasgow, UK, August 23 – 28, 2020, Proceedings, Part XIV 16. Springer International Publishing, 2020; 266 – 282.
- [39] GIDARIS S, BURSUC A, KOMODAKIS N, et al. Boosting few – shot visual learning with self – supervision [ C ] // Proceedings of the IEEE/CVF international conference on computer vision. 2019; 8059 – 8068.

## Relational embeddings and class covariance measures for few – shot image classification

ZHAO Rui, SHE Yu-mei, BAI Meng-ru

(School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650500, China)

**Abstract:** In the context of few – shot image classification, the structural and discriminative relationships between images have not been adequately emphasized, leading to imperfect feature embedding and unclear metric effects. To focus on the matching relationships between image features, a network structure based on relationship embedding and class covariance measurement is proposed. This is achieved by paying attention to the spatial information of individual images and matching features between images, realizing the relationship of metric embedding features. The proposed structure utilizes self – correlation attention and interactional correlation attention to learn the feature relationships within and between images. Specifically, the basic feature representation of individual images is mapped into self – correlation tensors, focusing on the spatial information of channels to learn structural information within image tensors. For image pairs, the mutual correlation information of the two image representations is calculated, generating shared attention to obtain discriminative relational information between images. Finally, the Mahalanobis distance is used to measure the feature relationships. After extensive experimental validation, the proposed method is evaluated on three few – shot benchmark datasets and compared with various existing methods. The results demonstrate that the network structure exhibits a certain effectiveness in learning the relationships between features.

**Key words:** few – shot learning; relationship embedding; correlation attention; mahalanobis distance

(责任编辑 段鹏)